Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Mutation Rates of STR Systems in Danes

Andersen, Kim Emil; Bøttcher, Susanne Gammelgaard; Christensen, Susanne; Eriksen, Poul Svante; Højbjerre, Malene; Hallenberg, C.; Simonsen, B.; Morling, N.

Publication date:
2007

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA):
Andersen, K. E., Bøttcher, S. G., Christensen, S., Eriksen, P. S., Højbjerre, M., Hallenberg, C., Simonsen, B., & Morling, N. (2007). *Mutation Rates of STR Systems in Danes*. Poster presented at International Society for Forensic Genetics (ISFG) Congress, Copenhagen, Denmark.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: May 05, 2023

Mutation Rates of STR Systems in Danes

K.E. Andersen*, S.G. Bøttcher*, E.S. Christensen*, P.S. Eriksen*, M. Højbjerre*, C. Hallenberg**, B. Simonsen**, N. Morling**
*Department of Mathematical Sciences, Aalborg University, Denmark
**Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health Sciences, University of Copenhagen, Denmark

Abstract

Danish paternity cases in the period 1999 to 2005 were investigated regarding mutation rates in STR loci. STR-typing was performed by the Applied Biosystems AmplfStr Profiler Plus kit in the period 1999 to early 2005, hereafter named the PP9, and by Applied Biosystems AmplfStr Identifier kit for the rest of 2005, hereafter named the IDFL. All cases with one to four genetic inconsistencies were manually inspected by two forensic geneticists and statistically analyzed by five statisticians. We found no significant effect of kits and no interaction of kits and STR loci, but differences in mutation rates on different STR loci. In the cases where mutations had occured, we found no interaction between kits, STR loci or sexes. However, we found differences in the mutation rates between the sexes, meaning that the differences in male and female mutation rates can be assumed constant over STR loci and kits. Sex and STR locus specific mutation rates were estimated with 95% confidence limits by the method of Clopper and Pearson (1934).

Data

Only cases, where consistency could be obtained by assuming at most one mutation, were included and considered as having the true father (PI $\geq$ 10,000).

Test for Hardy Weinberg equilibrium revealed the presence of two subpopulations - one defined by paternity cases in courts in the geographic area of Denmark and another one originating from Greenland. Further investigations were performed without the data from Greenland.

For PP9, 4,239 cases were analyzed at 9 STR loci and for IDFL, 1,244 cases were analyzed at 15 STR loci, in total 5,483 cases.

For each case, we considered the following four options:

- **Case consistent:** No inconsistency was found (hidden mutation was not discovered).
- **Paternal mutation:** An inconsistency was found and determined as a paternal mutation.
- **Indeterminable mutation:** An inconsistency was found, but it was not possible to determine whether it was a paternal or maternal mutation.
- **Maternal mutation:** An inconsistency was found and determined as a maternal mutation.

Further, in cases with an indeterminable mutation, if the possible maternal mutation was a one step mutation and the possible paternal mutation was a two (or more) step mutation, it was counted as a maternal mutation and vice versa.

In cases of an inconsistency between a homozygous parent and a homozygous child, null alleles were assumed.

All the inconsistent cases were manually checked. Cases where the inconsistency could be explained by a one step mutation, were counted as such. In the remaining cases, the original file was inspected and the problem was identified.

- Error in coding - corrected.
- A third allele not registered in the database - corrected.
- Insecurity about allele type - corrected.
- Father not present, so grandparents typed instead - left out of the analysis.
- Two step mutation - counted as such.

The manual inspection reduced the number of inconsistent cases and we emphasize the importance of scrutinizing the data because the estimation of rare events like mutations rates are very sensitive to small deviations.

The number of mutations at different STR loci for the three sexes (paternal, indeterminable or maternal) and for the two kits (PP9 or IDFL) are shown in Table 1 and 2.

Table 1 - PP9

STR Locus	Sex			Total
	Paternal	Indeterminable	Maternal	
D13S317	6	2	3	11
D18S51	10	4	3	17
D21S11	9	4	6	19
D3S1358	8	3	1	12
D5S818	6	3	1	10
D7S820	9	1	3	13
D8S1179	2	3	1	6
FGA	14	6	5	25
vWA	12	4	5	21
Total	76	30	28	134

Table 2 - IDFL

STR Locus	Sex			Total
	Paternal	Indeterminable	Maternal	
CSF1PO	1	0	0	1
D13S317	2	0	0	2
D16S539	0	0	1	1
D18S51	4	0	0	4
D19S433	1	0	0	1
D21S11	3	0	2	5
D2S1338	2	1	0	3
D3S1358	1	2	0	3
D5S818	0	2	0	2
D7S820	1	1	0	2
D8S1179	1	1	2	4
FGA	3	0	0	3
TH01	0	0	0	0
TPOX	0	1	0	1
vWA	1	1	0	2
Total	20	9	5	34

Statistical Methods

As the mutation rates are expected to be small, i.e. distributions asymmetrical, we have used non-approximate methods, e.g. Fisher's exact test for association between categorical variables. Estimates of rare occurences are sensitive to errors, therefore, scrutinizing of these kinds of data is of the outermost importance.

We investigated the possible associations between mutations, STR loci, amplification kits and sexes, which are all categorical variables. For the cases where a mutation had occured, i.e. given mutation, the association between sex, kit and STR loci was investigated in the framework of a graphical model (Lauritzen, 1996). By Fisher's exact test, we looked for association between sex and STR locus conditional on the kit. The next test was for association between kit and STR locus, followed by a test for independence between kit and sex.

For the investigation of possible effects on registered mutations rates of kit and STR locus (method induced mutation), we used a logistic regression model of mutation using kit, STR locus and interaction between them as independent variables.

The probability of a given mutation being paternal is significant different from being maternal as the 95 % confidence interval for male mutations probability did not contain the value 0.5.

Let x_{ij} denote the number of mutations for sex i = paternal, indeterminable or maternal, at STR locus j = D13S317, ..., vWA. Summation over index is indicated by $*$, e.g. $x_{*\,j} = \sum_i x_{ij}$. The total number of cases is denoted n. The STR locus specific mutation rate (per meiosis) was estimated by

$$\hat{p}_{\text{STR locus}} = \frac{x_{*\ \text{STR locus}}}{2n}$$

with corresponding confidence limits using the method by Clopper and Pearson (1934). The paternal mutation rates for a specific STR locus was estimated by

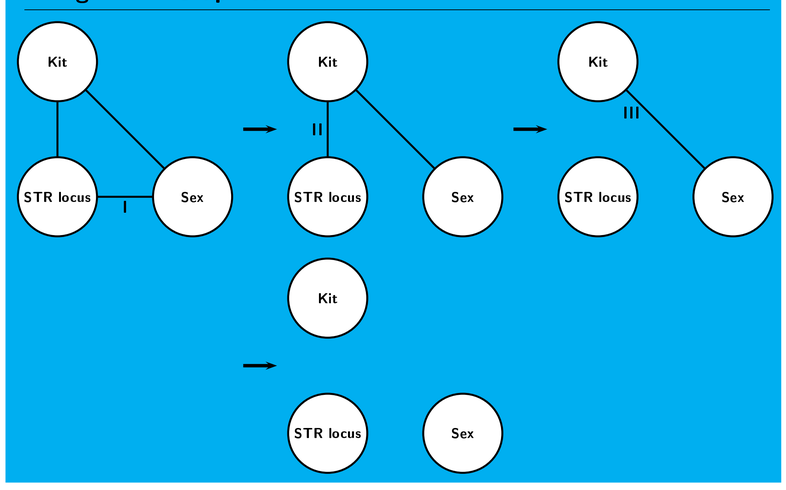
$$\hat{p}_{\text{paternal,STR locus}} = 2 \bullet \frac{x_{*\ \text{STR locus}}}{2n} \bullet \frac{x_{\text{paternal}\ *}}{x_{\text{paternal}\ *} + x_{\text{maternal}\ *} + x_{\text{indeterminable}\ *}}$$

Estimates for maternal and indeterminable mutation rates were made similarly.

Results

Initially, the data were analysed for the purpose of determining the conditional independency structures of the variables: STR locus, sex and kit given that mutation had occured. Starting with a graphical model assuming mutual dependencies of all variables (see Figure 1 - first graph), we reduced the model, i.e. remove edges, by performing serveral Fisher's exact tests for independence in marginal tabels. Non-approximate tests were necessary due to the relative small number of mutations. The three edges were tested in the order illustrated in Figure 1:

Figure 1 - Graphical Model



I: Test of independence of STR locus and sex conditioned on kit:

$$H_0: \quad \text{STR locus} \perp\!\!\!\perp \text{sex} \mid \text{kit}$$

corresponded to the tests of independences in Table 1 and 2. The H_0's were not rejected (p= 0.995 for PP9 and p = 0.119 for IDFL), i.e. the edge was removed.

II: Test of independence of STR locus and kit:

$$H_0: \quad \text{STR locus} \perp\!\!\!\perp \text{kit}$$

corresponded to the test of independence in Table 3 (combination of total columns in Table 1 and 2 with common STR locus). The H_0 was not rejected (p=0.6457), i.e. the edge was removed.

Table 3

Kit	STR Locus									Total
	D13S317	D18S51	D21S11	D3S1358	D5S818	D7S820	D8S1179	FGA	vWA	
PP9	11	17	19	12	10	13	6	25	21	134
IDFL	2	4	5	3	2	2	4	3	2	27
Total	13	21	24	15	12	15	10	28	23	161

III: Test of independence of sex and kit:

$$H_0: \quad \text{sex} \perp\!\!\!\perp \text{kit}$$

corresponded to the test of independence in Table 4 (combinination of total rows of Table 1 and 2). The H_0 was not rejected (p=0.7005), i.e. the edge was removed.

Table 4

Kit	Sex			Total
	Paternal	Indeterminable	Maternal	
PP9	76	30	28	134
IDFL	20	9	5	34
Total	96	39	33	168

In cases with mutations: STR locus, kit and sex were all mutually independent.

Results - cont'd

The analyses were followed by a logistic regression with mutation as dependent variable and STR locus, kit and a mixture of these as independent variables. We found no significant mixed effect of STR locus and the kit (p=0.6770) and no significant effect of the kit (p=0.0672), whereas a significant effect of STR locus (p=0.0256) was found.

We tested the probability of a given mutation being equally distributed in males and females. The probability of a given mutation being paternal was estimated from the total row in Table 4 as $\hat{p} = \frac{96}{96+33} = 0.744$ with the 95 % confidence interval as $[0.660; 0.817]$ (in terms of odds: $\frac{\hat{p}}{1-\hat{p}} = 2.91$ with the 95 % confidence interval as $[1.94; 4.46]$). Since $p = \frac{1}{2}$ is not included in the 95 % confidence interval, the probabilities are not equally distributed in male and female.

By aggregating over kit and sex for common STR locus, we observed the following numbers of mutations shown in Table 5 (which corresponds to the total row in Table 3) and Table 6.

Table 5 - PP9 & IDFL

STR Locus									Total
D13S317	D18S51	D21S11	D3S1358	D5S818	D7S820	D8S1179	FGA	vWA	
13	21	24	15	12	15	10	28	23	161

Table 6 - IDFL

STR Locus						Total
CSF1PO	D16S539	D19S433	D2S1338	THO1	TPOX	
1	1	1	3	0	1	7

By use of the number of mutations, we estimated the STR locus specific mutation rates with 95% confidence intervals computed by the method of Clopper and Pearson (1934). Table 7 states the mutation rates (in %) for common STR locus, and Table 8 states for IDFL specific STR loci. To estimate the rates, we used $n = 5,483$ in Table 7 and $n = 1,244$ in Table 8.

Table 7 - STR Locus Specific Mutation Rates - PP9 & IDFL

STR Locus	D13S317	D18S51	D21S11	D3S1358	D5S818	D7S820	D8S1179	FGA	vWA
Lower	0.06	0.12	0.14	0.08	0.06	0.08	0.04	0.17	0.13
Estimate	0.12	0.19	0.22	0.14	0.11	0.14	0.09	0.26	0.21
Upper	0.20	0.29	0.33	0.23	0.19	0.23	0.17	0.37	0.31

Table 8 - STR Locus Specific Mutation Rates - IDFL

STR Locus	CSF1PO	D16S539	D19S433	D2S1338	THO1	TPOX
Lower	0.00	0.00	0.00	0.02	0.00	0.00
Estimate	0.04	0.04	0.04	0.12	0.00	0.04
Upper	0.22	0.22	0.22	0.35	0.15	0.22

The paternal STR locus specific mutation rates (in %) are listed in Table 9. In case of indeterminable, the rates in Table 7 and 8 shall be multiplied by the probability for indeterminable, which was estimated by $\hat{p} = \frac{39}{96+39+33} = 0.232$, see Table 9. In case we are interested in the maternal STR locus mutations rates, the rates shall be multiplied by $\hat{p} = \frac{33}{96+33} = 0.2558$, see Table 9.

Table 9 - STR Locus and Sex Specific Mutation rates

STR Locus	D13S317	D18S51	D21S11	D3S1358	D5S818	D7S820	D8S1179	FGA	vWA
Paternal	0.14	0.22	0.25	0.16	0.13	0.16	0.10	0.29	0.24
Indetermine	0.06	0.09	0.10	0.06	0.05	0.06	0.04	0.12	0.10
Maternal	0.05	0.08	0.09	0.05	0.04	0.05	0.04	0.10	0.08
STR Locus	**CSF1PO**	**D16S539**	**D19S433**	**D2S1338**	**THO1**	**TPOX**			
Paternal	0.05	0.05	0.18	0.14	0.00	0.05			
Indetermine	0.02	0.02	0.07	0.06	0.00	0.02			
Maternal	0.02	0.02	0.06	0.05	0.00	0.02			

Discussion

- **Data cleansing:** We urge data cleansing, since estimates of rare occurences are very sensitive to errors of any kind, e.g. clerical errors in the registration of data.
- **Validation:** The data were highly validated by forensic geneticists and thoroughly analyzed by statisticians, implying high confidence to the mutation rates estimated.
- **Kit dependence:** The logistic regression showed a borderline, though not significant, effect of the kits. As the mutation rates are mainly based on cases from PP9, a follow-up study based on IDFL should be performed to investigate if the mutation rates of PP9 are in accordance with those of IDFL.

References

Clopper, C. J. and Pearson, E. S. (1934), The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial, *Biometrika* **26**, 404–413.

Lauritzen, S. L. (1996), *Graphical Models*, Clarendon Press, Oxford, UK.

Contact: E. Susanne Christensen, Department of Mathematical Sciences, Aalborg University, **e-mail:** susanne@math.aau.dk